Semi- and non-parametric competing risks analysis of right censored data and failure cause missing completely at random

Laurent Bordes, Jean-Yves Dauxois, Pierre Joly

Semi- and non-parametric competing risks analysis of right censored data and failure cause missing completely at random

Laurent Bordes
Université de Pau et des Pays de l'Adour - France
Laboratoire de Mathématiques Appliquées
UMR CNRS 5142

Jean-Yves Dauxois
Université de Franche-Comté - France
Département de Mathématiques
UMR CNRS 6623

Pierre Joly[1,2]
[1]Université Victor Segalen Bordeaux 2 - ISPED Bordeaux - France
[2]INSERM U875 Bordeaux - France

Abstract

We consider a nonparametric and a semiparametric (in presence of covariates) additive hazards rate competing risks model with censoring and failure cause possibly missing completely at random. Estimators of the unknown parameters are proposed in order to satisfy some optimality criteria. Large sample results are given for all the estimators. Our nonparametric method is applied to a real data set and the behavior of the semiparametric estimators are analyzed through a Monte Carlo study.
Key words: Additive hazards, competing risks, counting processes, missing failure cause, reliability, survival analysis.

1 Introduction

We consider $p \geq 2$ independent competing failure causes. We assume that to each failure time T_j is associated a hazard rate function (risk function) λ_j with $1 \leq j \leq p$. The failure time T is the minimum of the p failure times associated to the p failure causes (this can be seen as a 1-out-of-p system in reliability), then we have $T = T_1 \wedge \cdots \wedge T_p$. The failure time T can be censored by a censoring time C, then we observe $X = T \wedge C$ and $\delta = 1(T \leq C)$ where $1(\cdot)$ is the indicator function. Generally, when T is uncensored, that is for $\delta = 1$, the failure cause is known, which means that $\sum_{j=1}^{p} j1(T_j = T)$ is observed, but from times to times, it may happen that the failure cause is unknown and no partial information about the failure cause is available. In addition, a vector of explanatory variables denoted by Z and having potentially significant effects on the p failure times may be observed. In this paper we propose a model that allows to analyze such lifetime data from a semiparametric point of view in presence of covariates, and from a nonparametric point of view otherwise. We need to emphasize that our model extends some existing models in the semi-/non-parametric direction, however, the missingness mechanism is accounted here in the simplest way.

The problem of competing risks is not new and during the last two decades many models have been proposed in order to account that a system or an individual may fail or dead from several causes (see Crowder, 2001, for a large overview on the topic). In a number of real applications of competing risks models the authors have to face the problem of missing information (e.g. Miyakawa, 1984; Usher and Hodgson, 1988; Lin *et al.*, 1993; Schabe, 1994;

Goetghebeur and Ryan, 1995; Guttman *et al.*, 1995; Reiser *et al.*, 1996; Basu *et al.*, 1999; Flehinger *et al.*, 2002; Craiu and Duchesne, 2004; Craiu and Reiser, 2007).

Among the great amount of paper dealing with competing risks model some of them focus on nonparametric estimation methods (see e.g. Lo, 1991 and Schabe, 1994). Because in many case partial information about the failure cause can only be obtained (e.g. masked cause of failure), a large number of works developed some specific methods with accurate modeling of the missingness mechanism. Most of these models are parametric and when a latent variable represents the missingness mechanism an EM-type algorithm can be proposed to estimate the model parameters. In Craiu and Duchesne (2004) such estimation procedure is proposed and the missingness may depend both on the failure cause and the failure time. Recently, Craiu and Reiser (2007) considered a very complete parametric model including dependence of failure causes.

Some authors developed estimation procedures in the semi-/non-parametric framework for two or more failure causes (see e.g. Myakawa, 1984; Dinse, 1986; Lo, 1991, Schabe, 1994). The special case of a possibly censored single failure cause differs from the competing risks model only be the fact that in this case the censoring time is not an event of interest. However, when the censoring information is missing, we are close to the competing risks situation where failure causes are possibly missing. Some specific methods has been derived in Gijbels *et al.* (1993), McKeague and Subramanian (1998), van der Laan and McKeague (1998), Sun and Zhou (2003) and Subramanian (2004) for various models including or not covariates, and several missingness mechanisms.

Goetghebeur and Ryan (1995) proposed a competing risks model with proportional hazards assumption for the different failure causes. In their model the mechanism of missingness may depend on the failure time (this is the missing at random assumption) while in the model we propose it is independent of everything (this is the missing completely at random assumption). In our model each failure type has its own semiparametric additive hazards rate model and at the contrary to Goetghebeur and Ryan (1995), these failure rates are not linked.

The paper is organized as follows. In Section 2 we describe the model and we point out that each data resulting from the model can be seen as the realization of a nonhomogeneous Markov process. In Section 3 the estimators are defined. Because data for which the failure cause is missing are informative for the whole parameters of the model, we develop a method that allows to account this information in an optimal way with respect to an efficiency criterium. In Section 4, for the Euclidean parameters, and in Section 5 for the functional parameters, the corresponding estimators are shown to be consistent and asymptotically Gaussian. For each estimator, a consistent estimator of the asymptotic variance is provided. Section 6 is devoted to numerical examples. A Monte Carlo study is performed for the case including covariates whereas our estimation method is applied to a real data set that does not include covariates. Some concluding remarks are given in the last Section.

2 A semiparametric additive hazards model

The complete data are as follows.

- T_j is the duration associated to the jth cause ($j \in \{1, \ldots, p\}$).
- The law of T_j, conditional on the vector of covariates $Z \in \mathbb{R}^k$, is defined by the hazard rate function
$$\lambda_j(t|Z) = \lambda_{0j}(t) + \beta_j^T Z, \quad t \geq 0,$$
where λ_{0j} is the baseline hazard rate function of the jth cause, and $\beta_j \in \mathbb{R}^k$ is the regression parameter associated to the jth cause.

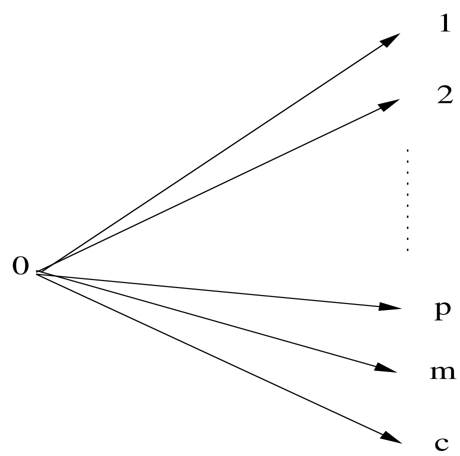

Figure 1: Markov graph associated to (X, M, D, Z)

- The censoring variable C either has a hazard rate function λ_C or is deterministic; it is independent of the T_j's conditionally on Z.
- The binary variable $\delta = 1(T \leq C)$, where $T = T_1 \wedge \cdots \wedge T_p$, is the censoring indicator.
- The binary variable M is equal to 1 when the failure cause is known (thus is is equal to 0 when $\delta = 0$).
- The variable $D = \delta M \sum_{j=1}^p j1(T_j \leq T)$ reveals the failure cause when the failure time is uncensored and $M = 1$.

The observation coming from one indiviual is therefore (X, δ, D, Z) where X is the observed duration. Denoting

$$P(M = 1|X, Z, \delta = 1) = P(M = 1|\delta = 1) = \alpha \in [0, 1]$$

and

$$P(M = 0|X, Z, \delta = 0) = P(M = 0|\delta = 0) = 1,$$

the observation (X, δ, D), conditional on Z, may be seen as the realization of a $(p+3)$-state nonhomogeneous Markov process (see Fig. 1). Writing $\bar{\lambda}_{0x}(\cdot|Z)$ the transition rate, conditional on Z, for the transition $0 \to x$ $(x \in \{1, \dots, p, m, c\})$ we obtain:

$$\left\{ \begin{array}{lcl} \bar{\lambda}_{0j}(t) & = & \alpha\lambda_j(t|Z) \text{ for } j \in \{1, \dots, p\}, \\ \bar{\lambda}_{0m}(t) & = & (1-\alpha)\sum_{j=1}^p \lambda_j(t|Z), \\ \bar{\lambda}_{0c}(t) & = & \lambda_C(t). \end{array} \right.$$

The independence of $T_1, \dots, T_p, C$, conditional on Z, is a sufficient condition to obtain the above transition rates.

Remark Up to the parameter α (or $1-\alpha$), the transition rates (excepted for the transition $0 \to c$) are additive hazards functions. This is obvious for the transition $\bar{\lambda}_{0j}$ with $1 \leq j \leq p$ but it remains true for the transition $0 \to m$, indeed:

$$\bar{\lambda}_{0m}(t) = (1-\alpha)\left(\lambda_m(t) + \beta_m^T Z\right),$$

where $\lambda_{0m} = \sum_{j=1}^p \lambda_{0j}$ et $\beta_m = \sum_{j=1}^p \beta_j$.

3 Regression parameters estimation

3.1 Data and notations

We denote by $(X_i, \delta_i, D_i, Z_i)_{1\leq i\leq n}$ n independent and identically distributed copies of (X, δ, D, Z).

For $j \in \{1, \dots, p, m\}$ we define the counting processes:

$$\begin{aligned} N_{ij}(t) &= 1(X_i \leq t, D_i = j) \text{ for } j \neq m, \\ N_{im}(t) &= 1(X_i \leq t, \delta_i = 1, D_i = 0). \end{aligned}$$

Hereafter, in order to simplify our notations, we write $m \equiv p+1$. Let Y_i be the risk process defined by $Y_i(t) = 1(X_i \geq t)$. Then, by Andersen *et al.* (1993) or Fleming and Harrington (1991), for $1 \leq i \leq n$ and $j \in \{1, \dots, p+1\}$ the M_{ij} processes defined by

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_i(s)\bar{\lambda}_{0j}(s)ds, \quad t \geq 0,$$

are $\mathbb{F}$-martingales with respect to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ defined by

$$\mathcal{F}_t = \sigma\{N_{ij}(s), Y_i(s); s \leq t; 1 \leq i \leq n, j \in \{1, \dots, p+1\}\}.$$

3.2 Estimation of the Euclidean parameters

We denote by $\tau < +\infty$ the upper bound of the interval of study. It means that individuals are observed on the time interval $[0, \tau]$. The particular case where the censoring variable C is deterministic corresponds to a Type-I censored sampling plan with $C = \tau$. With the assumption made on M the parameter α is naturally estimated by

$$\hat{\alpha} = \hat{\alpha}(\tau) = \frac{\sum_{i=1}^n 1(D_i > 0; X_i \leq \tau)}{\sum_{i=1}^n 1(\delta_i = 1; X_i \leq \tau)} = \frac{\sum_{j=1}^p N_{\cdot j}(\tau)}{N_{\cdot\cdot}(\tau)},$$

where any point in the subscripts corresponds to summation over all the possible values of the subscript. The estimator $\hat{\alpha}$ of α is natural since it reveals, among the individuals doing an uncensored transition, the proportion of individuals having a known cause of failure. Individuals doing the transition $0 \rightarrow j$ $(1 \leq j \leq p)$ allow to estimate β_j. Indeed, it is straightforward to extend the estimating function of Lin et Ying (1994). Therefore β_j is estimated by $\hat{\beta}_j$ which is the solution of $\mathcal{U}_j(\beta, \hat{\alpha}, \tau) = 0$ where

$$\mathcal{U}_j(\beta, \hat{\alpha}, \tau) = \sum_{i=1}^n \int_0^\tau \left[Z_i - \bar{Z}(s)\right] \left[dN_{ij}(s) - \hat{\alpha}\beta^T Z_i Y_i(s)ds\right],$$

with

$$\bar{Z}(s) = \frac{\sum_{i=1}^n Y_i(s) Z_i}{\sum_{i=1}^n Y_i(s)}.$$

We ever saw that the transition rate $0 \rightarrow m$ follows an *additive risk model*; using data from this transition allows to estimate β_m by $\hat{\beta}_m$ solution of $\mathcal{U}_m(\beta, \hat{\alpha}, \tau) = 0$ where

$$\mathcal{U}_m(\beta, \hat{\alpha}, \tau) = \sum_{i=1}^n \int_0^\tau \left[Z_i - \bar{Z}(s)\right] \left[dN_{im}(s) - (1-\hat{\alpha})\beta^T Z_i Y_i(s)ds\right].$$

At this stage, each β_j has its estimator $\hat{\beta}_j$ for $j = 1, \dots, p, p+1$. Because $\hat{\beta}_m$ is an estimator of $\beta_m = \beta_1 + \dots + \beta_p$, it is natural to use this available information in order to improve estimates of the β_j's. We adopt the following strategy. First we consider the block matrix H defined by

$$H = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1p+1} \\ H_{21} & H_{22} & \cdots & H_{2p+1} \\ \vdots & \vdots & & \vdots \\ H_{p1} & H_{p2} & \cdots & H_{pp+1} \end{pmatrix},$$

where for $1 \leq i \leq p$ and $1 \leq j \leq p+1$ H_{ij} is a $k \times k$-real valued matrix. The matrix H has to satisfy

$$H \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \beta_m \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

for all $1 \leq i \leq p$ and $\beta_i \in \mathbb{R}^k$. Moreover, we want H to minimize the function $\hat{q}(H)$ defined by

$$\hat{q}(H) = \mathrm{trace}(H\hat{\Sigma}H^T),$$

where $\hat{\Sigma}$ is an estimator of the asymptotic variance matrix of $(\hat{\beta}_1^T, \ldots, \hat{\beta}_p^T, \hat{\beta}_m^T)^T$. In other words, we are looking for a linear transformation of $(\hat{\beta}_1^T, \ldots, \hat{\beta}_p^T, \hat{\beta}_m^T)^T$ that remains an estimator of the β_j's for $j = 1, \ldots, p$ but that has a smaller variance (in some particular sense) than the initial one. Thus, denoting by $\hat{H}$ the matrix minimizing $\hat{q}(H)$, we denote by $\tilde{\beta}_i = \sum_{j=1}^{p+1} \hat{H}_{ij}\hat{\beta}_j$ (with $\hat{\beta}_{p+1} \equiv \hat{\beta}_m$) the final estimator of β_i for $i = 1, \ldots, p$. Later we refer this estimators as "minimal variance estimators" (in the above sense).

Remark It is worth to note that constraints imposed to H does not link its lines. Indeed, denoting by I_k the identity matrix of order k, these constraints may be written

$$\begin{cases} H_{ii} + H_{ip+1} & = & I_k \\ H_{ij} + H_{ip+1} & = & 0 \end{cases}$$

for $1 \leq i \leq p$ et $j \in \{1, \ldots, p\} \backslash \{i\}$.

On the other hand we have

$$\hat{q}(H) = \sum_{i=1}^{p} \mathrm{trace}(H_{i\bullet}\hat{\Sigma}H_{i\bullet}^T) = \sum_{i=1}^{p} \hat{q}_i(H)$$

where $\hat{q}_i(H) = \mathrm{trace}(H_{i\bullet}\Sigma H_{i\bullet}^T)$ and $H_{i\bullet}$ is the ith line block of H. Because the constraints are linear and does not link the lines of H it is sufficient to solve for each $i \in \{1, \ldots, p\}$ the following problem (P_i):
Looking for matrices $H_{i1}, \ldots, H_{ip+1}$ satisfying:

$$(P_i) \quad \begin{cases} H_{ii} + H_{ip+1} = I_k, \\ H_{ij} + H_{ip+1} = 0 \text{ for } j \neq i, \\ \mathrm{trace}(H_{i\bullet}\hat{\Sigma}H_{i\bullet}^T) \text{ is minimal.} \end{cases}$$

3.3 Example for $p = k = 2$

We have to solve problems (P_1) and (P_2) which are identical. As a consequence we consider only (P_1). Let us write $H^{(j)} = H_{1j}$ for $j = 1, 2, 3$ and introduce the following notations:

$$H^{(j)} = \begin{pmatrix} h_{11}^{(j)} & h_{12}^{(j)} \\ h_{21}^{(j)} & h_{22}^{(j)} \end{pmatrix} \quad \text{and} \quad \hat{Q} = \begin{pmatrix} \hat{\Sigma} & 0 \\ 0 & \hat{\Sigma} \end{pmatrix}$$

and $L = \left(h_{11}^{(1)}, h_{12}^{(1)}, h_{11}^{(2)}, h_{12}^{(2)}, h_{11}^{(3)}, h_{12}^{(3)}, h_{21}^{(1)}, h_{22}^{(1)}, h_{21}^{2)}, h_{22}^{(2)}, h_{21}^{(3)}, h_{22}^{(3)}\right) = (l_1, \ldots, l_{12})$.

Then

$$\hat{q}_1(H) = \operatorname{trace}\left(\left(H^{(1)}\ H^{(2)}\ H^{(3)}\right)\hat{\Sigma}\left(H^{(1)}\ H^{(2)}\ H^{(3)}\right)^T\right) = L\hat{Q}L^T.$$

The constraints are

$$\left\{\begin{array}{lcl} H^{(1)} + H^{(3)} & = & I_2 \\ H^{(2)} + H^{(3)} & = & 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} h_{ij}^{(1)} + h_{ij}^{(3)} & = & 1 \text{ for } 1 \le i,j \le 2, \\ h_{ij}^{(2)} + h_{ij}^{(3)} & = & 0 \text{ for } 1 \le i,j \le 2. \end{array}\right. \Leftrightarrow CL = d,$$

where

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad d = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Let λ be the Lagrange parameter associated to the optimization problem with linear constraints. The Lagrange function ℓ can be written

$$\ell(L, \lambda) = \frac{1}{2}L^T\hat{Q}L + (CL - d)^T\lambda.$$

At the optimum $\hat{L}$ it satisfies

$$\left\{\begin{array}{lclcl} \dfrac{\partial \ell}{\partial L}(\hat{L}, \lambda) & = & 0 & = & \hat{Q}\hat{L} + C^T\lambda, \\ C\hat{L} & = & d, & & \end{array}\right.$$

so we obtain

$$\lambda = -[C\hat{Q}^{-1}C^T]^{-1}d,$$

and hence

$$\hat{L} = \hat{Q}^{-1}C^T[C\hat{Q}^{-1}C^T]^{-1}d.$$

Remark Note that $\hat{Q}$ is invertible whenever $\hat{\Sigma}$ is.

4 Asymptotic behavior of the regression parameters estimators

4.1 Asymptotics for the $\hat{\beta}_j$

First we need some notations. Let z be a column vector in $\mathbb{R}^k$, we note

$$x^{\otimes l} = \left\{\begin{array}{lll} 1 & \text{if} & l = 0, \\ z & \text{if} & l = 1, \\ zz^T & \text{if} & l = 2. \end{array}\right.$$

For $0 \le l \le 2$ and $0 \le s \le \tau$ we define

$$S_l(s) = \frac{1}{n}\sum_{i=1}^{n} Y_i(s) Z_i^{\otimes l},$$

and for $b \in \mathbb{R}^k$ we define

$$S_3(s;b) = \frac{1}{n}\sum_{i=1}^{n} Y_i(s) Z_i^{\otimes 2} b^T Z_i.$$

Hereafter asymptotic results are given with respect to n tending to infinity. Let us introduce the following assumptions.

A1. The probability α of a known cause is strictly positive.

A2. The upper bound of the study interval satisfies $0 < \int_0^\tau \lambda_{0j}(s)ds < +\infty$ for $1 \leq j \leq p$ and the vectors of covariates Z_i are uniformly bounded with respect to $i \geq 1$.

A3. For $0 \leq l \leq 2$ there exist functions s_k defined on $[0, \tau]$ such that

$$\max_{0 \leq l \leq 2} \sup_{s \in [0,\tau]} \| S_l(s) - s_l(s) \| \xrightarrow{P} 0.$$

Moreover, s_0 is bounded by above by a strictly positive real number.

A4. Denoting by $a(u) = s_2(u) - s_1^{\otimes 2}(u)/s_0(u)$, the matrix

$$A(\tau) = \int_0^\tau a(u)du$$

is positive definite, and

$$\theta(\tau) = \int_0^\tau \left[s_0(u)\lambda_{0p+1}(u)du + \beta_{p+1}^T s_1(u) \right] du > 0.$$

A5. Let us consider $b \in \mathbb{R}^k$ and $S_4(s;b) = n^{-1}\sum_{i=1}^n (b^T Z_i) Y_i(s)(s_1(s) Z_i^T / s_0(s))$. For all $b \in \mathbb{R}^k$, there exist functions $s_3(s;b)$ and $s_4(s;b)$ such that

$$\max_{3 \leq l \leq 4} \sup_{s \in [0,\tau]} \| S_l(s;b) - s_l(s;b) \| \xrightarrow{P} 0.$$

A6. Functions a, $a\lambda_{0j}$, $s_0\lambda_{0j}$, s_1, $s_1^{\otimes 2}/s_0^2$, $s_3(\cdot;\beta_j)$ and $s_4(\cdot;\beta_j)$ (for $1 \leq j \leq p$) are integrable on $[0, \tau]$.

Remark Assumption A1 is clearly necessary to obtain the identifiability of the whole model parameters. Assumption A2 insures that data can be observed everywhere on $[0, \tau]$ for each failure cause in an homogeneous way; the fact that the Z_i are uniformly bounded could be relaxed, and they could also be time dependent but it would add some unimportant technicalities with respect to the main objective of this work. Assumption A3 is a simple consequence of the strong law of large numbers for i.i.d. (independent and identically distributed) processes. It is satisfied in particular when the covariates are i.i.d. with third order moment. Assumption A4 is necessary to identify β_j using the transition $0 \to j$ $(1 \leq j \leq p+1)$, whereas A5 and A6 are technical conditions.

Now we introduce the vectors

$$\hat\beta = (\hat\beta_1^T, \ldots, \hat\beta_{p+1}^T)^T \quad \text{and} \quad \beta = (\beta_1^T, \ldots, \beta_{p+1}^T)^T.$$

Theorem 4.1 *Under Assumptions A1-A5, the random vector* $\sqrt{n}(\hat\beta - \beta)$ *is asymptotically gaussian, centered, with positive definite covariance matrix* $\Sigma(\tau)$ *defined hereafter.*

Proof. First we denote by

$$\mathcal{U}_j(b,\alpha_j,\tau) = \sum_{i=1}^n \int_0^\tau \left(Z_i - \bar{Z}(s)\right)\left(dN_{ij}(s) - \alpha_j b^T Z_i Y_i(s) ds\right),$$

for $1 \leq j \leq p+1$ with $b \in \mathbb{R}^k$ and

$$\alpha_j = \begin{cases} \alpha & \text{if} \quad 1 \leq j \leq p, \\ 1-\alpha & \text{if} \quad j = p+1. \end{cases}$$

If we note

$$\hat{\alpha}_j = \begin{cases} \hat{\alpha} & \text{if} \quad 1 \leq j \leq p, \\ 1-\hat{\alpha} & \text{if} \quad j = p+1, \end{cases}$$

we have $\mathcal{U}_j(\hat{\beta}_j,\hat{\alpha}_j,\tau) = 0$ for $1 \leq j \leq p$. First let us remark that for $1 \leq j \leq p+1$ the following equality holds

$$\frac{1}{\sqrt{n}}\mathcal{U}_j(\beta_j,\alpha_j,\tau) = \hat{A}(\tau)\left[\hat{\beta}_j\sqrt{n}(\hat{\alpha}_j - \alpha_j) + \alpha_j\sqrt{n}(\hat{\beta}_j - \beta_j)\right], \tag{1}$$

where

$$\hat{A}(\tau) = \int_0^\tau \left[S_2(s) - S_1^{\otimes 2}(s)/S_0(s)\right] ds.$$

Now, using the definition of $\hat{\alpha}$ we have

$$\begin{aligned}
\hat{\alpha}(\tau) &= \frac{\displaystyle\sum_{i=1}^n \sum_{j=1}^p \int_0^\tau dN_{ij}(s)}{\displaystyle\sum_{i=1}^n \sum_{j=1}^{p+1} \int_0^\tau dN_{ij}(s)} \\
&= \frac{\displaystyle\sum_{i=1}^n \sum_{j=1}^p \int_0^\tau dM_{ij}(s) + \sum_{i=1}^n \sum_{j=1}^p \int_0^\tau Y_i(s)\bar{\lambda}_{0j}(s|Z_i) ds}{\displaystyle\sum_{i=1}^n \sum_{j=1}^{p+1} \int_0^\tau dM_{ij}(s) + \sum_{i=1}^n \sum_{j=1}^{p+1} \int_0^\tau Y_i(s)\bar{\lambda}_{0j}(s|Z_i) ds} \\
&= \frac{\displaystyle\sum_{j=1}^p M_{\cdot j}(\tau) + \alpha \sum_{i=1}^n \sum_{j=1}^p \int_0^\tau Y_i(s)\lambda_j(s|Z_i) ds}{\displaystyle M_{\cdot\cdot}(\tau) + \sum_{i=1}^n \sum_{j=1}^p \int_0^\tau Y_i(s)\lambda_j(s|Z_i) ds} \\
&= \frac{\displaystyle\sum_{j=1}^p M_{\cdot j}(\tau) + \alpha n \int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right] du}{\displaystyle M_{\cdot\cdot}(\tau) + n \int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right] du}.
\end{aligned}$$

It follows that

$$\begin{aligned}\sqrt{n}(\hat{\alpha}(\tau)-\alpha) &= \frac{\frac{1-\alpha}{\sqrt{n}}\sum_{j=1}^{p} M_{\cdot j}(\tau) - \frac{\alpha}{\sqrt{n}} M_{\cdot p+1}(\tau)}{\frac{1}{n}M_{\cdot\cdot}(\tau) + \int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right] du} 1(N_{\cdot\cdot}(\tau) > 0) \\ &\quad -\sqrt{n}\alpha 1(N_{\cdot\cdot}(\tau) = 0) \\ &= \frac{\frac{1-\alpha}{\sqrt{n}}\sum_{j=1}^{p} M_{\cdot j}(\tau) - \frac{\alpha}{\sqrt{n}} M_{\cdot p+1}(\tau)}{\frac{1}{n}M_{\cdot\cdot}(\tau) + \int_0^\tau \left[S_0(u)\lambda_{0p+1}(u) + \beta_{p+1}^T S_1(u)\right] du} + o_P(1).\end{aligned}$$

Using the Lenglart inequality we show that

$$\frac{1}{n}M_{\cdot\cdot}(\tau) \xrightarrow{P} 0.$$

Therefore, A3 and A4 yield

$$\sqrt{n}(\hat{\alpha}(\tau)-\alpha) = \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} \frac{1}{\sqrt{n}} M_{\cdot j}(\tau) + o_P(1), \tag{2}$$

where $\alpha_j^* = 1-\alpha$ for $1 \le j \le p$ and $\alpha_{p+1}^* = -\alpha$. By applying the Rebolledo theorem and using (2) it is easy to prove that $\hat{\alpha}$ is $\sqrt{n}$-consistent.

Because the processes $\mathcal{U}_j(\beta_j, \alpha_j, \cdot)$ are $\mathcal{F}_t$-martingales, and proving by using the Lenglart inequality that divided by n these processes converge to 0 uniformly in probability on $[0,\tau]$, we obtain the convergence in probability of $\hat{\beta}$ to β by using additionally the consistency of $\hat{\alpha}$, equality (1), and Assumptions A3 and A4. Once this convergence is established, we obtain by (1) the equality

$$\begin{aligned}\sqrt{n}(\hat{\beta}_j - \beta_j) &= \frac{1}{\alpha_j}\left(A^{-1}(\tau)\frac{\mathcal{U}_j(\beta_j,\alpha_j,\tau)}{\sqrt{n}} - \beta_j\sqrt{n}(\hat{\alpha}_j - \alpha_j)\right) \\ &\quad + \frac{1}{\alpha_j}\left(\left(\hat{A}^{-1}(\tau) - A^{-1}(\tau)\right)\frac{\mathcal{U}_j(\beta_j,\alpha_j,\tau)}{\sqrt{n}} - \sqrt{n}(\hat{\alpha}_j-\alpha_j)(\hat{\beta}_j - \beta_j)\right),\end{aligned}$$

where the second term of the right hand side is an $o_P(1)$ since $A_n^{-1}(\tau)$ converges in probability to $A^{-1}(\tau)$ from Assumptions A3 and A4, $\mathcal{U}_j(\beta_j,\alpha_j,\tau)/\sqrt{n}$ is a $O_P(1)$ (because it is asymptotically gaussian by the Rebolledo theorem), $\hat{\alpha}$ is $\sqrt{n}$-consistent and $\hat{\beta}$ converges to β. Finally the following approximation is obtained

$$\sqrt{n}(\hat{\beta}_j - \beta_j) = \frac{1}{\alpha}\left(A^{-1}(\tau)\frac{\mathcal{U}_j(\beta_j,\alpha_j,\tau)}{\sqrt{n}} - \beta_j\sqrt{n}(\hat{\alpha}-\alpha)\right) + o_P(1), \tag{3}$$

for $1 \le j \le p$, and

$$\begin{aligned}&\sqrt{n}(\hat{\beta}_{p+1} - \beta_{p+1}) \\ &= \frac{1}{1-\alpha}\left(A^{-1}(\tau)\frac{\mathcal{U}_{p+1}(\beta_{p+1},\alpha_{p+1},\tau)}{\sqrt{n}} + \beta_{p+1}\sqrt{n}(\hat{\alpha}-\alpha)\right) + o_P(1).\end{aligned} \tag{4}$$

Hence, by (2), (3), and (4) we obtain

$$\sqrt{n}(\hat{\beta} - \beta) = \Sigma_1(\tau)\mathcal{U}_n(\tau) + o_P(1), \tag{5}$$

where

$$\Sigma_1(\tau) = \begin{pmatrix} \frac{A^{-1}(\tau)}{\alpha} & 0 & \cdots & 0 & -\frac{\beta_1}{\alpha} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{A^{-1}(\tau)}{\alpha} & 0 & -\frac{\beta_p}{\alpha} \\ 0 & \cdots & 0 & \frac{A^{-1}(\tau)}{1-\alpha} & \frac{\beta_{p+1}}{1-\alpha} \end{pmatrix},$$

and

$$\mathcal{U}_n(\tau) = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathcal{U}_1(\beta_1, \alpha_1, \tau) \\ \vdots \\ \mathcal{U}_{p+1}(\beta_{p+1}, \alpha_{p+1}, \tau) \\ \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta(\tau)} M_{\cdot j}(\tau) \end{pmatrix}.$$

As a consequence, study the asymptotic behavior of $\sqrt{n}(\hat{\beta}-\beta)$ leads to study the asymptotic behavior of the process $\mathcal{U}_n$ which is an $\mathcal{F}_t$-martingale. As in Andersen and Gill (1982), we apply the Rebolledo theorem (see Rebolledo, 1980, and an adapted version in Andersen *et al.*, 1993, p. 83–84). This theorem allows to show that $\mathcal{U}_n$ converges weakly in $(D[0,\tau])^{k(p+1)+1}$ to a gaussian martingale. Thus, the random vector $\mathcal{U}_n(\tau)$ is asymptotically gaussian and centered. Because this theorem is now of classical use we only derive the limit of the predictible variation process of $\mathcal{U}_n$ that gives the asymptotic variance matrix of $\mathcal{U}_n(\tau)$. On one hand we have for $1 \leq j \leq p+1$

$$\begin{aligned} & \left\langle n^{-1/2}\mathcal{U}_j(\beta_j, \alpha, \cdot)\right\rangle(t) \\ = \quad & \alpha_j \int_0^t \left[S_2(s) - S_1^{\otimes 2}(s)/S_0(s)\right] \lambda_{0j}(s) ds \\ + \quad & \alpha_j \int_0^t \left[S_3(s;\beta_j) - S_4(s;\beta_j) - S_4^T(s;\beta_j) + (\beta_j^T S_1(s)) S_1^{\otimes 2}(s)/S_0^2(s)\right] ds + o_P(1), \end{aligned}$$

which by Assumptions A2-A6 converges to

$$\Theta_j(t) =$$

$$\alpha_j \int_0^t \left[a(s)\lambda_{0j}(s) + s_3(s;\beta_j)\beta_j - s_4(s;\beta_j) - s_4^T(s;\beta_j) + (\beta_j^T s_1(s)) s_1^{\otimes 2}(s)/s_0^2(s)\right] ds.$$

On the other hand, for $1 \leq j \neq j' \leq p+1$, we have

$$\langle \mathcal{U}_j(\cdot), \mathcal{U}_{j'}(\cdot)\rangle(t) = 0,$$

where we note $\mathcal{U}_j(t) = \mathcal{U}_j(\beta_j, \alpha_j, t)$ for $1 \leq j \leq p+1$. Thus the asymptotic covariance of terms $\mathcal{U}_j$ and $\mathcal{U}_{j'}$ is null. It remains to calculate the asymptotic covariance of $\sqrt{n}(\hat{\alpha}-\alpha)$ with the $n^{-1/2}\mathcal{U}_j$'s, and with itself. It is easy to show that for $1 \leq k \leq p+1$ we have

$$\frac{1}{n}\left\langle \mathcal{U}_k(\cdot), \sum_{j=1}^{p+1} \frac{\alpha_j^* M_{\cdot j}(\cdot)}{\theta(\tau)}\right\rangle(t) \xrightarrow{P} \kappa_k \frac{\alpha(1-\alpha)}{\theta(\tau)} A(t)\beta_k \equiv \xi_k(t),$$

where $\kappa_k = 1$ if $1 \leq k \leq p$ and $\kappa_{p+1} = -1$, and

$$\frac{1}{n}\left\langle \sum_{j=1}^{p+1} \frac{\alpha_j^* M_{\cdot j}(\cdot)}{\theta(\tau)}\right\rangle(t) \xrightarrow{P} \frac{\alpha(1-\alpha)}{\theta(\tau)} \equiv \xi_\alpha(t).$$

Then, setting

$$\Sigma_2(\tau) = \begin{pmatrix} \Theta_1(\tau) & 0 & 0 & \xi_1(\tau) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \Theta_{p+1}(\tau) & \xi_{p+1}(\tau) \\ \xi_1^T(\tau) & \cdots & \xi_{p+1}^T(\tau) & \xi_\alpha(\tau) \end{pmatrix}$$

we obtain by (5) that $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically gaussian with asymptotic covariance matrix $\Sigma(\tau) = \Sigma_1(\tau)\Sigma_2(\tau)\Sigma_1^T(\tau)$. The matrix $\Sigma_2(\tau)$ being positive definite, the same holds for $\Sigma(\tau)$. □

4.2 The optimal estimator

Giving the asymptotic behavior of $\tilde{\beta}$ requires to propose an estimator $\hat{\Sigma}$ of the asymptotic covariance matrix $\Sigma(\tau)$. In this paragraph, when it is possible, we avoid to indicate dependence in τ, for exemple $\Sigma(\tau)$ will be denoted by Σ.

It is easy to check that for $1 \le j \le p+1$ we have

$$\hat{\beta}_j = \frac{1}{\hat{\alpha}_j}\hat{A}^{-1}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left(Z_i - \bar{Z}(s)\right)dN_{ij}(s)$$

Then, for $1 \le j \le p+1$ we note

$$\hat{\Theta}_j = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left(Z_i - \bar{Z}(s)\right)^{\otimes 2}dN_{ij}(s),$$

and

$$\hat{\xi}_j = \kappa_j\frac{\hat{\alpha}(1-\hat{\alpha})}{\hat{\theta}}\hat{A}\hat{\beta}_j,$$

where $\hat{\theta} = N_{\cdot\cdot}(\tau)/n$, and finally

$$\hat{\xi}_\alpha = \frac{\hat{\alpha}(1-\hat{\alpha})}{\hat{\theta}}.$$

Therefore we set

$$\hat{\Sigma}_1 = \begin{pmatrix} \frac{\hat{A}^{-1}}{\hat{\alpha}} & 0 & \cdots & 0 & -\frac{\hat{\beta}_1}{\hat{\alpha}} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{\hat{A}^{-1}}{\hat{\alpha}} & 0 & -\frac{\hat{\beta}_p}{\hat{\alpha}} \\ 0 & \cdots & 0 & \frac{\hat{A}^{-1}}{1-\hat{\alpha}} & \frac{\hat{\beta}_{p+1}}{1-\hat{\alpha}} \end{pmatrix}$$

and

$$\hat{\Sigma}_2 = \begin{pmatrix} \hat{\Theta}_1 & 0 & 0 & \hat{\xi}_1 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \hat{\Theta}_{p+1} & \hat{\xi}_{p+1} \\ \hat{\xi}_1^T & \cdots & \hat{\xi}_{p+1}^T & \hat{\xi}_\alpha \end{pmatrix}.$$

Let $\hat{H}$ be the matrix defined at Section 3, $\tilde{\beta} = \hat{H}\hat{\beta}$ and $\beta^* = (\beta_1^T, \ldots, \beta_p^T)^T$.

Theorem 4.2 *Under Assumptions A1-A5, the matrix $\hat{\Sigma} = \hat{\Sigma}_1\hat{\Sigma}_2\hat{\Sigma}_1^T$ converges in probability to the matrix Σ. If Σ is invertible, then $\sqrt{n}(\tilde{\beta} - \beta^*)$ has an asymptotically centered gaussian distribution, whose the covariance matrix trace is less or equal to trace($H\Sigma H^T$) for all $H \in \mathcal{H}$. Such an estimator is said asymptotically T–optimal.*

Proof. The convergence of $\hat{A}$ to A is straightforward using Assumptions A3 and A4. Otherwise, we ever proved the convergence in probability of $\hat{\alpha}$ and $\hat{\beta}_j$ to α and β_j, respectively (for $1 \leq j \leq p+1$) in the proof of Theorem 4.1. As a consequence, we obtain the convergence in probability of $\hat{\Sigma}_1$ to Σ_1.

The convergence in probability of $\hat{\theta}$ to $\theta(\tau)$ is a consequence of the Lenglart inequality. These convergences yield the convergence in probability of the $\hat{\xi}_j$'s to the $\xi_j(\tau)$'s (for $1 \leq j \leq p+1$). Using repeatedly the Lenglart inequality allows to prove that $\hat{\xi}_\alpha$ and $\hat{\Theta}_j$ converge in probability to ξ_α and Θ_j, respectively (for $1 \leq j \leq p+1$). It follows that $\hat{\Sigma}_2$ converges in probability to Σ_2, and thus, that $\hat{\Sigma}$ converges in probability to Σ.

Finally, as we discussed in Sections 3.2 and 3.3, the matrix $\hat{H}$ depends continuously on $\hat{\Sigma}$, then when n tends to infinity, $\hat{H}$ converges in probability to the matrix $H_\Sigma \in \mathcal{H}$ that minimizes trace$(H\Sigma H^T)$ whose the existency follows from the positive definiteness of Σ. Moreover, because $\hat{H}\beta = \beta^*$, by the Slutsky lemma $\sqrt{n}(\tilde{\beta} - \beta^*) = \hat{H}\sqrt{n}(\hat{\beta} - \beta)$ has a centered asymptotic gaussian distribution whose the covariance matrix $H_\Sigma \Sigma H_\Sigma^T$ satisfies trace$(H_\Sigma \Sigma H_\Sigma^T) \leq$ trace$(H\Sigma H^T)$ for all H such that $H\beta = \beta^*$. This finishes the proof. $\square$

5 Estimation of functional parameters

5.1 Without covariate: nonparameteric case

This is the simplest case but the model is interesting for application. Indeed the unknown parameters of the model are the failure rates λ_j associated to each failure cause. The transition rates of the Markov graph (see Fig. 1) satisfy for $t \geq 0$,

$$\begin{cases} \bar{\lambda}_{0j}(t) &= \alpha\lambda_j(t) \text{ pour } j \in \{1, \ldots, p\}, \\ \bar{\lambda}_{0m}(t) &= (1-\alpha)\sum_{j=1}^p \lambda_j(t) \equiv (1-\alpha)\lambda_m(t), \\ \bar{\lambda}_{0c}(t) &= \lambda_C(t). \end{cases}$$

At the contrary to the model by Goetghebeur and Ryan (1995), when there is no covariate in the data, the model is still of interest because it allows different failure rates for failure causes.

5.1.1 Principle of an optimal estimator of the cumulative hazard rate functions

Let Λ_j be the cumulative hazard rate function associated to the jth failure cause ($1 \leq j \leq p$). By standard arguments it is natural to estimate Λ_j by

$$\hat{\Lambda}_j(t) = \frac{1}{\hat{\alpha}} \sum_{i=1}^n \int_0^t \frac{dN_{ij}(s)}{Y(s)}$$

and $\Lambda_m(t) = \sum_{j=1}^p \Lambda_j(t)$ by

$$\hat{\Lambda}_m(t) = \frac{1}{1-\hat{\alpha}} \sum_{i=1}^n \int_0^t \frac{dN_{im}(s)}{Y(s)},$$

where $Y(s) = \sum_{i=1}^n 1(X_i \geq s)$ and $\hat{\alpha}$ is defined in Section 3. Again we have an estimator of each Λ_j ($1 \leq j \leq p$) and an estimator of their sum. Let us denote by $\hat{H}(t)$ the $p \times (p+1)$ time dependent matrix, that will be specified later, and $\hat{\Lambda}(t) = (\hat{\Lambda}_1(t), \ldots, \hat{\Lambda}_p(t), \hat{\Lambda}_m(t))^T$ one estimator of $\Lambda(t) = (\Lambda_1(t), \ldots, \Lambda_p(t), \Lambda_m(t))^T$. We define $\tilde{\Lambda}(t) = \hat{H}(t)\hat{\Lambda}(t)$ as one estimator of $\Lambda^*(t) = (\Lambda_1(t), \ldots, \Lambda_p(t))^T$. If the family of matrices $\{\hat{H}(t); t \in [0,\tau]\}$ satisfies $\hat{H}(t)a = a^*$ for all $t \in [0,\tau]$, $a^* = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$ and $a = (a^{*T}, \sum_{j=1}^p a_j)^T$. Then if $\hat{H}(t)$

converges in probability to $H(t)$ and if the sequence of processes $(\sqrt{n}(\hat{\Lambda}(t) - \Lambda(t)); t \in [0,\tau])$ converges weakly in $(D[0,\tau])^{p+1}$ to a centered gaussian process with covariance function $\Gamma(t)$, then the sequence of processes $(\sqrt{n}(\tilde{\Lambda}(t) - \Lambda^*(t)); t \in [0,\tau])$ converges weakly in $(D[0,\tau])^p$ to a gaussian process having variance function $H(t)\Gamma(t)H^T(t)$.

Let $\hat{\Gamma}(t)$ be a consistent estimator of $\Gamma(t)$, and $\hat{H}(t)$ be defined by

$$\hat{H}(t) = \arg\min_{H\in\mathcal{H}} \operatorname{trace}(H\hat{\Gamma}(t)H^T), \tag{6}$$

where $\mathcal{H}$ is the set of $p \times (p+1)$ matrices satisfying $Ha = a^*$ for all $a^* = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$ and $a = (a^{*T}, \sum_{j=1}^p a_j)^T$. We denote by L the column vector in $\mathbb{R}^{p(p+1)}$ defined by $L = (H_1, \ldots, H_p)$ where H_i is the ith line of $H \in \mathcal{H}$. The link between $L = (l_i)_{1\le i\le p(p+1)}$ and $H = (h_{ij})_{1\le i\le p; 1\le j\le p+1}$ is therefore $h_{ij} = l_{(i-1)(p+1)+j}$. Because linear constraints on H are transmitted to L we have

$$\begin{aligned}
&\left\{\begin{array}{lcl} h_{ii} + h_{i,p+1} &=& 1, \text{ for } 1 \le i \le p, \\ h_{ij} + h_{i,p+1} &=& 0, \text{ for } 1 \le i \le p,\ 1 \le j \le p+1 \text{ and } i \ne j, \end{array}\right. \\
\Leftrightarrow\ &\left\{\begin{array}{lcl} l_{(i-1)(p+1)+i} + l_{i(p+1)} &=& 1, \text{ for } 1 \le i \le p, \\ l_{(i-1)(p+1)+j} + l_{(i+1)p} &=& 0, \text{ for } 1 \le i \le p,\ 1 \le j \le p+1 \text{ and } i \ne j, \end{array}\right. \\
\Leftrightarrow\ & CL = d.
\end{aligned}$$

Denoting by $\hat{Q}(t)$ the block diagonal matrix defined by

$$\hat{Q}(t) = \left.\begin{pmatrix} \hat{\Gamma}(t) & & \\ & \ddots & \\ & & \hat{\Gamma}(t) \end{pmatrix}\right\} p\text{-times},$$

we show that $\hat{q}(L) \equiv \operatorname{trace}(H\hat{\Gamma}(t)H^T) = L^T\hat{Q}(t)L$. Following the method of Section 3.3, with the linear constraint $CL = d$, $\hat{q}(L)$ reaches it minimum at

$$\hat{L}(t) = \hat{Q}^{-1}(t)C^T(C\hat{Q}^{-1}(t)C^T)^{-1}d.$$

The $\hat{H}(t)$ matrix is therefore defined by $\hat{h}_{ij}(t) = \hat{l}_{(i-1)(p+1)+j}(t)$ for $1 \le i \le p$ and $1 \le j \le p+1$.

5.1.2 Asymptotics for $\hat{\Lambda}$ and $\tilde{\Lambda}$

Let us introduce two assumptions that allow to obtain the asymptotic behavior of estimators $\hat{\Lambda}$ and $\tilde{\Lambda}$.

B1. τ satisfies $0 < \Lambda_j(\tau) < +\infty$ for $1 \le j \le p$.

B2. There exists a function y, defined on $[0,\tau]$, and bounded away from 0, such that

$$\sup_{s\in[0,\tau]} |Y(s)/n - y(s)| \xrightarrow{P} 0.$$

Note that functions $y\lambda_j$ and λ_j/y are integrable on $[0,\tau]$ and we define

$$\eta_j(t) = \alpha_j \int_0^t \lambda_j(s)/y(s)ds$$

for $1 \le j \le m$ and $\eta_m = \sum_{j=1}^m \eta_j$. Assumption B2 is fulfilled whenever $\int_0^\tau \lambda_C(s)ds < +\infty$.

Theorem 5.1 *Under Assumption A1, B1, and B2, the sequence* $(\sqrt{n}(\hat{\Lambda}(t)-\Lambda(t)); t \in [0,\tau])$ *converges weakly on* $(D[0,\tau])^{p+1}$ *to a centered gaussian process with variance function* $\Gamma(t)$ *defined subsequently.*

Proof. The arguments we need here are close to those we used in the proof of Theorem 4.1, as a consequence we only give the main lines of this proof. We begin with $\hat{\alpha}$. Following the proof of Theorem 4.1, the equality (2) becomes

$$\sqrt{n}(\hat{\alpha}-\alpha) = \sum_{j=1}^{p+1} \frac{\alpha_j^*}{\theta_0(\tau)} \frac{1}{\sqrt{n}} M_{\cdot j}(\tau) + o_P(1), \tag{7}$$

where $\theta_0(\tau) = \int_0^\tau y(s)\lambda_m(s)ds > 0$ by B1 and B2. The Rebolledo theorem insures that $\hat{\alpha}(\tau)$ converges to α at $\sqrt{n}$-rate. With this result and the Lenglart inequality we show that the $\hat{\Lambda}_j$'s converge in probability to the Λ_j's, uniformly on $[0,\tau]$ for $1 \le j \le p+1$ $(p+1 \equiv m)$. In consequence we obtain the following uniform (in $t \in [0,\tau]$) approximation results

$$\sqrt{n}(\hat{\Lambda}_j(t)-\Lambda_j(t)) = -\frac{\Lambda_j(t)}{\alpha_j}\sqrt{n}(\hat{\alpha}_j-\alpha_j) + \frac{1}{\alpha_j}\sqrt{n}\int_0^t \frac{dM_{\cdot j}(s)}{Y(s)} + o_P(1) \tag{8}$$

for $1 \le j \le p+1$. By the approximations (7) and (8) we obtain

$$\sqrt{n}(\hat{\Lambda}(t)-\Lambda(t)) = \Sigma_3(t)\mathcal{V}_n(t) + o_P(1), \tag{9}$$

where

$$\Sigma_3(t) = \begin{pmatrix} \frac{1}{\alpha} & 0 & \cdots & 0 & -\frac{\Lambda_1(t)}{\alpha} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{1}{\alpha} & 0 & -\frac{\Lambda_p(t)}{\alpha} \\ 0 & \cdots & 0 & \frac{1}{1-\alpha} & \frac{\Lambda_m(t)}{1-\alpha} \end{pmatrix}$$

and

$$\mathcal{V}_n(t) = \sqrt{n}\begin{pmatrix} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)} \\ \vdots \\ \int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} \\ \frac{1}{n}\sum_{j=1}^{p+1}\frac{1-\alpha_j}{\theta_0(\tau)}M_{\cdot j}(\tau) \end{pmatrix}.$$

The Rebolledo theorem allows to prove that $\mathcal{V}_n$ converges weakly in $(D[0,\tau])^{p+2}$ to a centered gaussian process. It remains to specify the variance function Σ_4 of this process. First we can write $\mathcal{V}_n(t) = \mathcal{V}_{1,n}(t) + \mathcal{V}_{2,n}(t)$ where

$$\mathcal{V}_{1,n}(t) = \sqrt{n}\begin{pmatrix} \int_0^t \frac{dM_{\cdot 1}(s)}{Y(s)} \\ \vdots \\ \int_0^t \frac{dM_{\cdot p+1}(s)}{Y(s)} \\ \frac{1}{n}\sum_{j=1}^{p+1}\frac{1-\alpha_j}{\theta_0(\tau)}M_{\cdot j}(t) \end{pmatrix}$$

and

$$\mathcal{V}_{2,n}(t) = \frac{1}{\sqrt{n}}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sum_{j=1}^{p+1}\frac{1-\alpha_j}{\theta_0(\tau)}(M_{\cdot j}(\tau)-M_{\cdot j}(t)) \end{pmatrix}.$$

Because the $M_{\cdot j}$'s are martingales, we have

$$E\left[\mathcal{V}_n^{\otimes 2}(t)\right] = E\left[\mathcal{V}_{1,n}^{\otimes 2}(t)\right] + E\left[\mathcal{V}_{2,n}^{\otimes 2}(t)\right].$$

We are thus looking for

$$\Sigma_4(t) = \lim_{n\to\infty} E\left[\mathcal{V}_{1,n}^{\otimes 2}(t)\right] + \lim_{n\to\infty} E\left[\mathcal{V}_{2,n}^{\otimes 2}(t)\right].$$

Setting $\rho_j(t) = \kappa_j \alpha(1-\alpha)\Lambda_j(t)/\theta_0(\tau)$ for $1 \le j \le p+1 = m$ and $\rho_\alpha(\tau) = \alpha(1-\alpha)/\theta_0(\tau)$, it is straightforwardly established that

$$\Sigma_4(t) = \begin{pmatrix} \eta_1(t) & 0 & 0 & \rho_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \eta_{p+1}(t) & \rho_{p+1}(t) \\ \rho_1(t) & \cdots & \rho_{p+1}(t) & \rho_\alpha(\tau) \end{pmatrix}.$$

Then by (9) we obtain $\Gamma(t) = \Sigma_3(t)\Sigma_4(t)\Sigma_3^T(t)$, which finishes the proof. □

For $t \in]0,\tau]$ we define

$$\hat\Sigma_3(t) = \begin{pmatrix} \frac{1}{\hat\alpha} & 0 & \cdots & 0 & -\frac{\hat\Lambda_1(t)}{\hat\alpha} \\ 0 & \ddots & & \vdots & \vdots \\ \vdots & & \frac{1}{\hat\alpha} & 0 & -\frac{\hat\Lambda_p(t)}{\hat\alpha} \\ 0 & \cdots & 0 & \frac{1}{1-\hat\alpha} & \frac{\hat\Lambda_m(t)}{1-\hat\alpha} \end{pmatrix}$$

and

$$\hat\Sigma_4(t) = \begin{pmatrix} \hat\eta_1(t) & 0 & 0 & \hat\rho_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \hat\eta_{p+1}(t) & \hat\rho_{p+1}(t) \\ \hat\rho_1(t) & \cdots & \hat\rho_{p+1}(t) & \hat\rho_\alpha \end{pmatrix}$$

where for $1 \le j \le p+1$ we note

$$\hat\eta_j(t) = n\int_0^t \frac{dN_{\cdot j}(s)}{Y^2(s)},$$

and for $1 \le j \le p$ we note

$$\hat\rho_j(t) = \kappa_j \frac{\hat\alpha(1-\hat\alpha)\hat\Lambda_j(t)}{\hat\theta_0}$$

with $\hat\theta_0 = N_{\cdot\cdot}(\tau)/n$ and $\hat\rho_\alpha = \hat\alpha(1-\hat\alpha)/\hat\theta_0$. Thus we define $\hat\Gamma(t) = \hat\Sigma_3(t)\hat\Sigma_4(t)\hat\Sigma_3^T(t)$.

Theorem 5.2 *Under Assumptions A1, B1 and B2, for all $t \in [0,\tau]$, the matrix $\hat\Gamma(t)$ converges in probability to the matrix $\Gamma(t)$. Let us assume moreover that for all $t \in]0,\tau]$ the matrix $\Gamma(t)$ is invertible. If $\hat H(t)$ is the unique solution of (6), then $\sqrt{n}(\tilde\Lambda - \Lambda)$ converges weakly in $(D[0,\tau])^p$ to a centered gaussian process, and is asymptotically T-optimal with variance function equal to $H(t)\Gamma(t)H^T(t)$ (see (6)).*

Proof. Since the proof follows the lines of Theorem 4.2 proof, it is omitted. □
textitRemark Quantities $\hat\Lambda(t)$ and $\hat\Gamma(t)$ are piecewise constant, therefore it is sufficient to calculate the matrix $\hat H(t)$ at points $t \in [0,\tau]$ where $N_{\cdot\cdot}$ has jumps, that is at points $X_i \in [0,\tau]$ such that $\delta_i = 1$.

5.1.3 Survival functions estimation

One purpose of this paper is to improve the estimators of p parameters when in addition we are able to estimate their sum. Our model is especially well adapted to this problem for estimation of both regression parameters and cumulative hazard rate functions. Because the survival function coming from transition $0 \to m$ has not the additive property of the above mentioned parameters it is desirable to estimate the survival functions using the T-optimal estimates of the cumulative hazard functions. As a consequence, the Kaplan-Meier type estimator of S_j is defined by

$$\tilde{S}_j(t) = \prod_{s \le t} (1 - \Delta\tilde{\Lambda}_j(s)), \quad t \ge 0,$$

for $1 \le j \le p$, with $\Delta F(t) = F(t) - F(t-)$ for any $F : \mathbb{R} \to \mathbb{R}^k$. Because estimates $\tilde{\Lambda}_j$ are piecewise constant with jumps at points X_i for which $\delta_i = 1$, we have

$$\tilde{S}_j(t) = \prod_{\{1 \le i \le n; X_i \le t, \delta_i = 1\}} (1 - \Delta\tilde{\Lambda}_j(X_i)), \quad t \ge 0,$$

which are easily computed. Now let us define $1_p = (1, \ldots, 1)^T \in \mathbb{R}^p$ and for $t \in [0, \tau]$,

$$\tilde{S}(t) = (\tilde{S}_1(t), \ldots, \tilde{S}_p(t))^T \quad \text{and} \quad S(t) = (S_1(t), \ldots, S_p(t))^T.$$

Then we have

$$\tilde{S}(t) = \prod_{s \le t} \left[1_p - \Delta\tilde{\Lambda}(s)\right],$$

or equivalently, $\tilde{S} = \Phi(\tilde{\Lambda})$ where the operator Φ from $(D[0, \tau])^p$ to $(D[0, \tau])^p$ is defined from its p identical components product–integral ϕ from $D[0, \tau]$ to $D[0, \tau]$. As ϕ is Hadamard differentiable on $E_K \subset D[0, \tau]$ the set of functions with variations bounded by $K < +\infty$ (see Andersen *et al.*, 1993, p. 114) the same result holds for Φ on E_K^p with Hadamard derivative

$$(d\Phi(X).H)(t) = \Phi(X)(t) \odot H(t),$$

where for $a, b \in \mathbb{R}^p$, $a \odot b \in \mathbb{R}^p$ is the component-wise product of a and b. From Theorem II.8.2 in Andersen *et al.* (1993, p. 112), $\sqrt{n}(\tilde{S} - S)$ is asymptotically equivalent to $d\Phi(\Lambda) \odot \sqrt{n}(\tilde{\Lambda} - \Lambda^*) = S \odot \sqrt{n}(\tilde{\Lambda} - \Lambda^*)$. The next theorem follows from the preceding result and Theorem 5.2.

Theorem 5.3 *Under Assumptions A1, B1 and B2, the process* $(\sqrt{n}(\tilde{S}(t) - S(t)); t \in [0, \tau])$ *converges weakly in* $(D[0, \tau])^p$ *to a centered Gaussian process with covariance function* $\Omega(t) = (\omega_{ij}(t))_{1 \le i,j \le p}$ *defined by*

$$\omega_{ij}(t) = S_i(t)S_j(t)(H(t)\Gamma(t)H^T(t))_{ij}.$$

Then $\hat{\Omega}(t) = (\hat{\omega}_{ij}(t))_{1 \le i,j \le p}$ *defined by*

$$\hat{\omega}_{ij}(t) = \tilde{S}_i(t)\tilde{S}_j(t)(\hat{H}(t)\hat{\Gamma}(t)\hat{H}^T(t))_{ij},$$

converges in probability to $\Omega(t)$ *for* $t \in]0, \tau]$.

5.2 With explanatory variables

Let us consider $j \in \{1, \ldots, p\}$. For simplicity we only base our estimator of the conditional cumulative hazard function $\Lambda_j(\cdot|Z)$ for cause j, on transition $0 \to j$ (see Figure 1). Let us recall that $\Lambda_j(\cdot|Z)$ is defined by

$$\Lambda_j(t|Z) = \Lambda_{0j}(t) + \beta_j^T Z t, \quad t \ge 0.$$

The conditional survival function $S_j(\cdot|Z)$ associated to the jth cause is therefore defined by

$$S_j(t|Z) = \exp(-\beta_j^T Zt)\phi(\Lambda_{0j})(t), \quad t \geq 0,$$

where ϕ is still the product-integral operator.

First let us propose an estimator of Λ_{0j}. Using

$$dN_{\cdot j}(s) = dM_{\cdot j}(s) + \alpha S_0(s)\lambda_{0j}(s)ds + \alpha S_1^T(s)\beta_j ds,$$

we obtain the following estimator of $\Lambda_{0j}(t)$ defined by

$$\hat{\Lambda}_{0j}(t) = \frac{1}{\hat{\alpha}} \int_0^t \frac{dN_{\cdot j}(s)}{S_0(s)} - \hat{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)} ds, \quad t \geq 0,$$

where we recall that $\hat{\beta}_j$ satisfies $\mathcal{U}_j(\hat{\beta}_j, \alpha_j, \tau) = 0$ (see Section 3.2). Then we propose to estimate $\Lambda_j(\cdot|Z)$ by $\hat{\Lambda}_j(\cdot|Z)$ defined for $t \geq 0$ by

$$\begin{aligned} \hat{\Lambda}_j(t|Z) &= \frac{1}{\hat{\alpha}} \int_0^t \frac{dN_{\cdot j}(s)}{S_0(s)} + \hat{\beta}_j^T \left(tZ - \int_0^t \frac{S_1(s)}{S_0(s)} ds \right), \\ &= \sum_{\{1 \leq i \leq n; X_i \leq t, D_i = j\}} \frac{1}{\hat{\alpha} S_0(X_i)} + \hat{\beta}_j^T \left(tZ - \int_0^t \frac{S_1(s)}{S_0(s)} ds \right). \end{aligned}$$

The survival function $S_j(\cdot|Z)$ is therefore estimated by $\hat{S}_j(\cdot|Z)$ defined by

$$\hat{S}_j(t|Z) = \exp\left(-t\hat{\beta}_j^T Z + \hat{\beta}_j^T \int_0^t \frac{S_1(s)}{S_0(s)} ds \right) \prod_{1 \leq i \leq n; X_i \leq t, D_i = j} \left(1 - \frac{1}{\hat{\alpha} S_0(X_i)} \right).$$

Let $L_j(t)$ be the $(p+2) \times 1$-matrix and $\Xi_j(t)$ be the $(p+1) \times (p+2)$-matrix respectively defined by

$$L_j(t) = \frac{1}{\alpha} \begin{pmatrix} 1 \\ \Lambda_j(t|Z) - 2(k(t) + tZ)^T \beta_j \\ A^{-1}(\tau)(k(t) + tZ) \end{pmatrix},$$

and

$$\Xi_j(t) = \begin{pmatrix} \mu_{1j}(t) & \mu_{2j}(t) & \mu_{3j}^T(t) \\ \mu_{2j}(t) & \xi_\alpha(\tau) & \xi_j^T(\tau) \\ \mu_{3j}(t) & \xi_j(\tau) & \Theta_j(\tau) \end{pmatrix},$$

where $\xi_\alpha(\tau)$, $\xi_j(\tau)$ and $\Theta_j(\tau)$ are defined in Section 4 and for $1 \leq j \leq p$

$$k(t) = \int_0^t \frac{s_1(s)}{s_0(s)} ds,$$

$$\mu_{1j}(t) = \alpha \int_0^t \left(\frac{s_0(s)\lambda_{0j}(s) + \beta_j^T s_1(s)}{s_0^2(s)} \right) ds,$$

$$\mu_{2j}(t) = \frac{\alpha}{\theta(\tau)} \int_0^t \left(\frac{s_0(s)\lambda_{0j}(s) + \beta_j^T s_1(s)}{s_0(s)} \right) ds,$$

and

$$\mu_{3j}(t) = \alpha \int_0^t \frac{a(s)}{s_0(s)} ds \beta_j.$$

Let us define $\hat{L}_j$ and $\hat{\Xi}_j$ the empirical versions of L_j and Ξ_j, by

$$\hat{L}_j(t) = \frac{1}{\hat{\alpha}} \left(\begin{array}{c} 1 \\ \hat{\Lambda}_j(t|Z) - 2(\hat{K}(t) + tZ)^T \hat{\beta}_j \\ \hat{A}^{-1}(\tau)(\hat{K}(t) + tZ) \end{array} \right),$$

with

$$K(t) = \int_0^t \frac{S_1(s)}{S_0(s)} ds,$$

and

$$\hat{\Xi}_j(t) = \left(\begin{array}{ccc} \hat{\mu}_{1j}(t) & \hat{\mu}_{2j}(t) & \hat{\mu}_{3j}^T(t) \\ \hat{\mu}_{2j}(t) & \hat{\xi}_\alpha & \hat{\xi}_j^T \\ \hat{\mu}_{3j}(t) & \hat{\xi}_j & \hat{\Theta}_j \end{array} \right),$$

where $\hat{\xi}_\alpha$, $\hat{\xi}_j$ and $\hat{\Theta}_j$ are defined in Section 4 and

$$\hat{\mu}_{1j}(t) = n\hat{\alpha} \int_0^t \frac{dN_{\cdot j}}{S_0^2(s)},$$

$$\hat{\mu}_{2j}(t) = \frac{\hat{\alpha}}{\hat{\theta}} \int_0^t \frac{dN_{\cdot j}(s)}{S_0(s)},$$

and

$$\hat{\mu}_{3j}(t) = \hat{\alpha}\hat{A}(t)\hat{\beta}_j.$$

The proof of the next theorem is basically the same as the preceding ones then it is omitted.

Theorem 5.4 *Under Assumptions A1–A5 and if for all $t \in [0,\tau]$ the matrices $L(t)$ and $\Xi(t)$ are well defined, then for $1 \le j \le p$, the sequence of processes $(\sqrt{n}(\hat{\Lambda}_j(t|Z) - \Lambda_j(t|Z)); t \in [0,\tau])$ converges weakly in $D[0,\tau]$ to a centered Gaussian process with variance function γ_j defined by $\gamma_j(t) = L_j^T(t)\Xi_j(t)L_j(t)$. Moreover $\hat{\gamma}_j(t) = \hat{L}_j^T(t)\hat{\Xi}_j(t)\hat{L}_j(t)$ converges in probability to $\gamma_j(t)$ for all $t \in [0,\tau]$. The sequence of processes $(\sqrt{n}(\hat{S}_j(t|Z) - S_j(t|Z)); t \in [0,\tau])$ converges weakly in $D[0,\tau]$ to a centered Gaussian process with variance function π_j defined by $\pi_j(t) = S_j^2(t|Z)\gamma_j(t)$ and $\hat{\pi}_j(t) = \hat{S}_j^2(t|Z)\hat{\gamma}_j(t)$ converges in probability to $\pi_j(t)$ for all $t \in [0,\tau]$.*

6 Numerical study

6.1 A simulated example with covariates

In this section we give an example where we have two competing risks with a single covariate Z. We assume that Z is 1 or 2 with equal probabilities. Conditional on Z, we suppose that the hazard rate functions of T_1 and T_2 are respectively given by $\lambda_1(t|Z) = t/2 + 2Z$ and $\lambda_2(t|Z) = 2t/9 + 3Z$, therefore we have $\beta_1 = 2$ and $\beta_2 = 3$. The censoring time C is exponentially distributed with mean 1 whereas the rate of missingness α is equal to 0.5. As a consequence, on the whole simulated data the rate of censored data is about 12%, the rates of observed failures from cause 1 and 2 are respectively about 15% and 28%, and the rate of missingness is about 45%.

In Table 1 we compare the performances of estimators $(\hat{\beta}_i)_{i=1,2}$ based on transitions $0 \to i$ with T-optimal estimators $(\tilde{\beta}_i)_{i=1,2}$ that use observations coming from the three informative transitions. We computed the mean and standard errors of 1000 estimates of β_1 and β_2. We can see in Table 1 that from the bias point of view the T-optimal estimators are generally better (except for β_1 when $n = 100$) whereas the standard errors of T-optimal estimates are always smaller than the standard errors of estimates based on transitions 1 and 2.

n	$\hat{\beta}_1$	$\tilde{\beta}_1$	$\hat{\beta}_2$	$\tilde{\beta}_2$
100	2.116 [1.583]	2.117 [1.438]	3.024 [1.923]	3.004 [1.610]
200	2.055 [1.111]	2.023 [1.000]	3.083 [1.367]	3.039 [1.146]
400	2.021 [0.768]	2.001 [0.687]	3.043 [0.919]	3.011 [0.784]
1000	1.998 [0.500]	2.001 [0.454]	2.989 [0.584]	2.993 [0.502]

Table 1: Comparison of means and standard errors (within brackets) of $N = 1000$ estimates of $\hat{\beta}_i$ and $\tilde{\beta}_i$ for $i = 1, 2$.

6.2 A reliability example without covariates

This example deals with the hard drives data sample that may be found in Flehinger *et al.* (2002). These authors consider a scenario in which a company manufacturing hard drives for computers tries to analyze causes of failures of a certain sub-assembly. Some of these causes, such as "defective head", are related to components, but others (e.g., "particle contamination") are not; in this example, there are three major causes of failures which, without going into details, are denoted as causes 1, 2 and 3. We assume that these causes act independently and in series. 10,000 drives were manufactured and then information about failures was collected in a database during 4 years. The number of failures observed in this period was 172. Some of the failures were masked and a selected number of those were analyzed to complete resolution in the defect isolation laboratory. The only observed masked groups were {1, 2, 3} and {1, 3}. Considering causes 1 and 3 as a single failure cause we obtain data from a competing risks model with two failure modes. Mode 1 (corresponding to causes 1 , 3 or masked group {1, 3} of the original data set) and mode 2 (that corresponds to cause 2 in the original data set). The failure cause is missing when none of the three original cause of failure is known (corresponding to the masking group {1, 2, 3} of the original data set). Finally, we obtain a data set with 119 failures of type 1, 19 failures of type 2 and 34 failures for which the failure cause is unavailable. The lifetimes of the 9828 drives still functioning at the end of the study are censored by the 4 years of the study duration. Because no information is available after 4 years we fix τ to 4. The probability α that the two failure causes are missing is estimated by $\hat{\alpha} = 0.802$. Figure 2 shows, for each failure cause, estimates of cumulative hazard rate functions with and without using the transition $0 \to 3$ and the corresponding 95% pointwize confidence intervals for each estimate. We can see that there is a little gain to use transition $0 \to 3$. We can see also that the modified Nelson-Aalen estimators that we propose are slightly more regular (smaller size jumps) than the corresponding Nelson-Aalen estimators, in exchange of which our estimators can be non legitimate (they can be locally decreasing). However this drawback disapears as the sample size increases because of uniform convergence of our estimators. This is also true for for the two causes reliability estimates that are given on Figure 3 with pointwize 95% confidence intervals.

7 Conclusions

In this paper we consider a semiparametric competing risks model that accounts covariates and the fact that information on failure cause can be missing completely at random. In this paper the missingness mechanism is very simple because it is independent of everything (time, failure causes, covariates, etc.). However, because of nonparametric assumptions for baseline hazard rate function of every type of failure cause, this model is quite adaptable. It is certainly possible to extend this model in several directions: dependent risks, missingness mechanism dependent on failure cause, etc. This model can also be extended to the case of masked cause of failure for which many sophisticated parametric models and inference

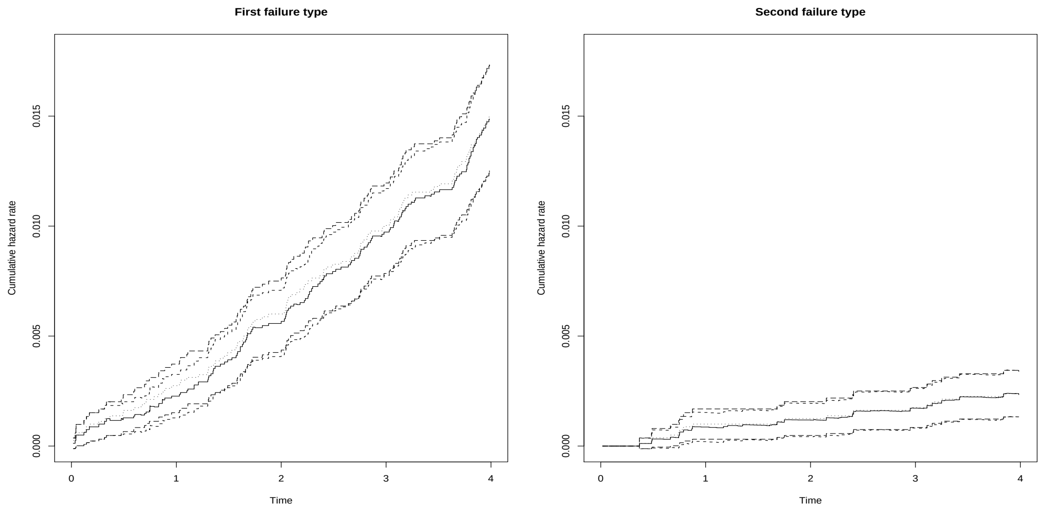

Figure 2: Cumulative hazard rate functions estimated without using the transition $0 \to 3$ (dotted lines) with 95% pointwize confidence intervals (long-dashed lines) and optimized estimation of the cumulative hazard functions (solid lines) with 95% pointwize confidence intervals (dashed lines) for the two failure causes.

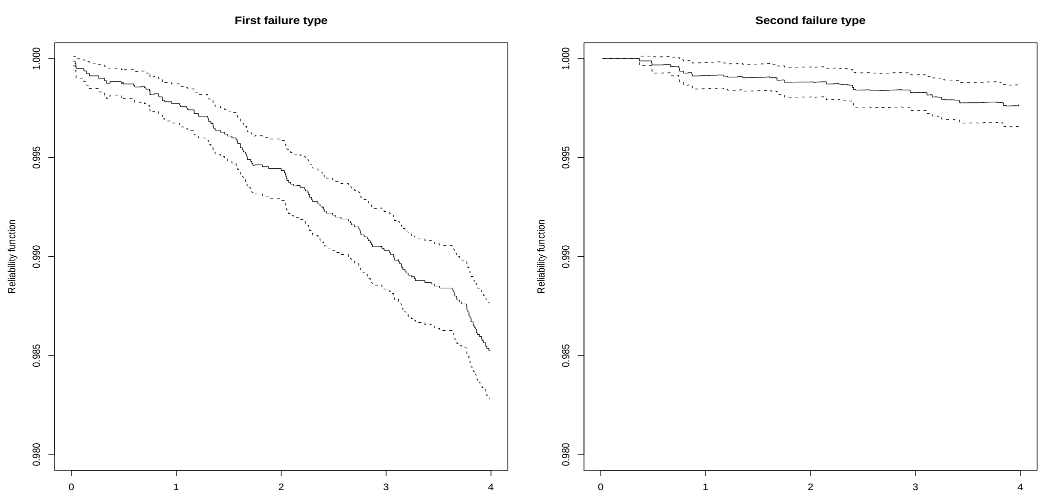

Figure 3: Estimated reliability functions (solid lines) with 95% pointwize confidence intervals (dashed lines) for the two failure causes.

methods have been developed over the two past decades. The estimation method that consists in seeing data as realizations of a nonhomogeneous markov process is inspired by McKeague and Subramanian (1998) while the estimators of the regression parameters of each transition is inspired from the Lin and Ying (1994) method. In addition we propose a linear transformation of these estimators which is shown to be asymptotically optimal in the sense of variance reduction.

References

[1] Andersen, P.K., Borgan, O. Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

[2] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study, *Ann. Statist.*, **10**, 1100–1120.

[3] Basu, S., Basu, A.P. and Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using non-identical Weibull models, *J. Statist. Plann. Inference*, **78**, 255–275

[4] Craiu, R.V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure, *Biometrika*, **91**, 543–558.

[5] Craiu, R.V. and Reiser, B. (2007). Inference for the dependent competing risks model with masked causes of failure, *Lifetime Data Analysis*, to appear.

[6] Crowder, M. (2001). *Classical Competing Risks*, Chapman & Hall/CRC Press, London.

[7] Dinse, G.E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data, *J. Amer. Statist. Assoc.*, **81**, 328–336.

[8] Flehinger, B.J., Reiser, R. and Yashchin, E. (2002). Parametric modeling for survival with competing risks and masked failure causes, *Lifetime Data Analysis*, **8**, 177–203.

[9] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*, Wiley, New York.

[10] Gijbels, I., Lin, D.Y. and Ying, Z. (1993). Non- and semi-parametric analysis of failure time data with missing failure indicators, *Preprint*.

[11] Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing, *Biometrika*, **82**, 821–833.

[12] Guttman, I., Lin, D.K.J., Reiser, B. and Usher, J.S. (1995). Dependent masking and system life data analysis: Bayesian inference for two-component systems, *Lifetime Data Analysis*, **1**, 87–100.

[13] Lin, D.K.J., Usher, J.S., Guess, F.M. (1993). Exact maximum likelihood estimation under masked system data, *IEEE Trans. Reliability*, **42**, 631–635.

[14] Lin, D.Y. and Ying, Z. (1994). Semiparametric estimation of the additive risk model, *Biometrika*, **81**, 61–71.

[15] Lo, S.-H. (1991). Estimating a survival function with incomplete cause-of-death data, *J. Multivariate Anal.*, **39**, 217–235.

[16] McKeague, I.W. and Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information, *Scandinavian Journal of Statistics*, **25**, 589–601.

[17] Miyakawa, M. (1984). Analysis of incomplete data in competing risks model, *IEEE Trans. Reliability*, **33**, 293–296.

[18] Rebolledo, R. (1980). Central limit theorems for local martingales, *Z. Wahrsch. verw. Geb.*, **51**, 269–286.

[19] Reiser, B., Flehinger, B.J., Conn, A.R. (1996). Estimating component-defect probability from masked system success/failure data, *IEEE Trans. Reliability*, **45**, 238–243.

[20] Subramanian, S. (2004). Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *J. Nonparametr. Stat.*, **16**, 797–817.

[21] Usher, J.S., Hodgson, T.J., (1988). Maximum likelihood analysis of component reliability using masked system life-test data, *IEEE Trans. Reliability*, **37**, 550–555.

[22] van der Laan, M. and McKeague, I.W. (1998). Efficient Estimation from Right-Censored Data when Failure Indicators are Missing at Random, *Ann. Statist.*, **26**, 164–182.

[23] Zhou, X. and Sun, L. (2003). Additive hazards regression with missing censoring information, *Statistica Sinica*, **13**, 1237–1257.

[24] Schabe, H. (1994). Nonparametric estimation of component lifetime based on masked system life test data, *J. Royal Statist. Soc. Ser. B* **56**, 251–259.

Laurent Bordes
Université de Pau et des Pays de l'Adour
Laboratoire de Mathématiques Appliquées
Bâtiment I.P.R.A
Avenue de l'Université - BP 1155
64013 PAU CEDEX - France
`laurent.bordes@univ-pau.fr`